DOCUMENT RESUME

ED 173 355 TM 008 315

AUTHOR Johnson, Dale D.; And Others

TITLE Format Effects on Vocabulary Assessment.

INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.

SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO WRDCIS-TR-474
PUB DATE Nov 78
CONTRACT OB-NIE-G-78-0217
NOTE 48p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cloze Procedure; Context Clues; Elementary Education; *Reading Ability; Reading Comprehension; Response Style (Tests); Semantics; Student Testing; *Test Construction; Test Items; *Verbal Tests; *Vocabulary
IDENTIFIERS *Item Format (Tests); Metropolitan Achievement Tests; *Synonyms

ABSTRACT

Three paradigms for assessing children's vocabulary knowledge were evaluated in this study: synonym, synonym in context, and cloze. The vocabulary tests, a paper-and-pencil adaptation of the on-line tests used in two previous studies, were administered to 830 second through seventh grade students. Subjects performed similarly on the three formats across the six grades tested. No one test format was better than either of the others within or across grades. But there were significant mean differences in scores obtained under the three formats overall. These differences ranked synonym, synonym in context, and cloze in order of difficulty from easiest to most difficult. When subjects were stratified by reading comprehension ability, scores on any one format were not significantly higher overall, than scores on either of the other two formats. Some differences did appear for the stratified groups. Low comprehenders scored higher on synonym than on synonym in context, and intermediate comprehenders scored somewhat higher on cloze than on synonym in context. It is possible that further differences in vocabulary performance might appear if children are grouped into finer categories of comprehension level. (Author/RD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED173355

Technical Report No. 474

FORMAT EFFECTS ON VOCABULARY ASSESSMENT

by Dale D. Johnson, Susan D. Pittelman,
Linda Shriberg, Judy Schwenker, and
Carol Morgan-Janty

November 1978

Wisconsin Research
and Development Center
for Individualized Schooling

IM008 315

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Rossmiller
Wisc R & D Ctr.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Technical Report No. 474

FORMAT EFFECTS ON VOCABULARY ASSESSMENT

by

Dale D. Johnson, Susan D. Pittelman, Linda Shriberg,
Judy Schwenker, and Carol Morgan-Janty

Report from the Project on
Studies in Language: Reading and Communication

Dale D. Johnson
Faculty Associate

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

November 1978

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Contract No. OB-NIE-G-78-0217

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

Acknowledgments

We gratefully appreciate the assistance of the teachers and pupils at Dean Elementary School, Algonquin Elementary School, and Brown Deer Middle School in Brown Deer, Wisconsin. We are especially appreciative of the assistance of Janet Juergens, District Reading Coordinator, for making the arrangements for testing in the schools.

Table of Contents

	Page
Acknowledgments	iv
List of Tables	vii
List of Figures	ix
Abstract	xi
Measurement of Vocabulary Knowledge	1
Method	5
Subjects	5
Stimuli	5
Procedure	7
Results	12
Summary	26
References	29
Appendix A	31

List of Tables

Table		Page
1	Subject Population by School and Grade	6
2	Subjects by Grade Level, Format and Level of Vocabulary Test, and Level of the Metropolitan Achievement Tests . .	9
3	Raw Score Summary Statistics on Vocabulary Tests	13
4	Pearson Correlations of Vocabulary Formats x Metropolitan Subtest Listed by Grade	14
5	Pearson Correlations of Vocabulary Formats x Metropolitan Subtest Across Grades	15
6	*t*-test Comparisons of Vocabulary Formats (Unstratified on Metropolitan Subtest)	16
7	One Way Anova of Main Effect of Format	18
8	Summary Statistics of Comprehension Ability Groups	19
9	Summary Statistics of Format Groups Stratified into Comprehension Ability Groups	20
10	*t*-tests of Unweighted Means of Factors of the Metropolitan Subtest and Vocabulary on Stratified Groups	22
11	Comparison of Comprehension Groups Within Formats of Vocabulary Measure	23
12	Comparison of Comprehension Ability Groupings Across Vocabulary Formats	24
13	Sequential Hypotheses Test Via Two Way Anova Under Fixed and Random Effects Models	25

List of Figures

Figure		Page
1	Sample items from level 4 of the vocabulary test	8

Abstract

This study is a continuation of two previous studies designed to evaluate techniques for assessing children's vocabulary knowledge. Three paradigms for assessing vocabulary knowledge were evaluated in this study: synonym, synonym in context, and cloze. The vocabulary tests used were a paper-and-pencil adaptation of the on-line tests used in the two previous studies.

Overall, subjects performed similarly on all three vocabulary formats and scores on all three formats correlated significantly with comprehension scores. When subjects' scores on the vocabulary tests were stratified by their performance on a standardized reading comprehension test, however, format effects did emerge. Whether more pronounced format effects would appear if subjects were grouped into finer categories of comprehension level is uncertain. Such elaborate stratification procedures were not believed to be justified at this point.

Certainly the question of format effects is an important one. If the particular format used to assess vocabulary causes subjects of varying comprehension abilities to perform differentially, the obtained scores must be interpreted accordingly. It may well be that no one paradigm for assessing vocabulary is most effective for all grades

and ability groups of students. Nevertheless, information is needed about specific testing modes, so that informed choices can be made about which of these modes is most appropriate for a particular testing purpose.

Measurement of Vocabulary Knowledge

The study described in this report is a continuation of two previous studies in which techniques for assessing vocabulary knowledge were assessed on-line, using a PLATO terminal. In the first study (Venezky, Perry, Chicone, and Pittelman, 1977) three techniques were assessed: self-screening, word matching under brief exposure, and synonyms. A follow-up study (Johnson, Pittelman, Schwenker, and Perry, 1978) examined the performance of elementary school children on five paradigms for assessing vocabulary knowledge on-line: synonym, synonym in context, cloze, oral recognition, and self-selection.

Due to the inaccessibility and high cost of PLATO terminals in most schools, the PLATO mode of assessment was discontinued after the completion of these two studies. The emphasis of the research was then changed from on-line diagnosis to evaluation through the use of a group administered, paper-and-pencil test.

The main purpose of this current study was to evaluate children's performance on three paradigms for assessing vocabulary knowledge: synonym, synonym in context, and cloze. The vocabulary tests used were a paper-and-pencil adaptation of the on-line tests used in the previous study, with modifications based on information from item analyses performed in the previous studies.

A child who demonstrates knowledge of a word on a particular format, is likely to show that he or she knows those same words when tested another way. It is, therefore, important to learn which of the many formats utilized for testing word knowledge correlate most highly with reading comprehension. Specifically, this study continued to examine the vocabulary test formats of synonym, synonym in context, and cloze, in order to analyze patterns of vocabulary test scores and to relate these formats to scores on a global measure of reading comprehension.

Most teachers value direct instruction of vocabulary words. The work of Davis in 1944, Spearitt in 1972, and others has shown that knowledge of word menaings is integral to reading comprehension.

What are words? Words may be referred to as morphemes, free morphemes, physical symbols for concepts, word-length units of meaning, graphic configurations bordered by space, or simply words. But, however labeled, they are inescapably important components of language, which *in their written forms*, must be dealt with by readers.

In a recent position paper on vocabulary development, Block (1976) cited Carroll's (1964) definition of concepts:

> Words often correspond to, or come to stand for, concepts. A word can be thought of as a physical symbol; the meaning of the word corresponds to those aspects of the concept that are shared by members of the same speech community; hence, a meaning of a word is a societally standardized concept. When we say that a word stands for or names a concept, it is implied that

> we are referring to the set of attributes that are commonly understood to be contained in the concept. Stated another way, the distinction is often made between a concept as a personal mental construct (*my* personal concept of the attributes of democracy; *your* personal concept of the attributes of democracy) and as a public entity (those attributes of democracy that are common to both of our concepts of democracy, and shared with other members of our speech community). (Block, 1976, pp. 43-44)

There are many ways of "knowing" a word. Humans have listening, speaking, reading, and writing vocabularies, which do not totally overlap. Some words are recognized aurally but not visually or are used in speaking but not in writing. Fluent reading obviously requires high-speed recognition of words and their syntactic arrangements. High frequency words, in particular, must be immediately identified by the fluent reader.

There has been a long history of attempts to form word lists which are important to learners, as evidenced by the more than 3,000 entries in the Bibliography of Vocabulary Studies (Dale, 1965). Basic word lists, derived from many sources, have been constructed for a number of purposes. Word lists have been compiled from such sources as the speech of young children, school essays and themes, language of bilingual adults, comic book words, computer analyses, award-winning children's literature, college profanities, and phonic irregularities. But most word lists have been constructed according to word frequencies.

Otto and Chester (1972) have stated their views on the major purpose of a sight word list:

> Most teachers of reading subscribe to the notion that beginning readers ought to be taught to recognize instantly at sight, a number of basic, frequently appearing words. The most convincing rationale for teaching sight words is that if they are well selected they will, because of their high frequency in printed materials, have high utility at all levels of reading development. (p. 435)

Sight word lists of frequently occurring words abound, and most instructional programs include their own vocabulary lists. This study does not purport to establish a priority list of words children need to be taught in order to become proficient readers. In practice, classroom teachers determine which words to teach and which meanings of those words to emphasize. Most often, teachers base these decisions on the materials they choose, the word lists they endorse, the instructional approaches they utilize, and the interests of their pupils. But, unfortunately, important diagnostic information is not available because the best ways of assessing children on word meanings are not yet known. An intensive literature review and ERIC search failed to locate even one study which attempted to analyze methods of testing word knowledge. The current study was undertaken to answer questions about vocabulary assessment.

Method

Subjects

A total of 830 students in grades two through seven participated in the study (see Table 1). The students were from two elementary schools and one middle school in Brown Deer, Wisconsin. The Brown Deer community has a middle and upper-middle class socioeconomic population. All subjects were from heterogeneous classrooms. Testing was conducted over a 3-day period in mid-January 1978. The testing session for each class lasted approximately 30 minutes.

Stimuli

The vocabulary test used was a group administered, paper-and-pencil test designed to examine three paradigms for assessing vocabulary knowledge: synonym, synonym in context, and cloze. For each paradigm, there were 8 graduated levels of the test with 10 items per level. Each level of the test was color-coded for easy identification. Target words and response words for all tests were derived from the PLATO version of the vocabulary test. These words have been based on a selection of words appearing both in standardized vocabulary tests and in The Living Word Vocabulary (Dale & O'Rourke, 1976).

The Synonym Test required that subjects choose, from among four response words, the word closest in meaning to a target word. The Synonym in Context Test presented the target word within the context of a sentence. Subjects were to choose, from among four response words, the word closest in meaning to the target word. The Cloze Test required

Table 1

Subject Population by

School and Grade

Brown Deer School District	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Total per School
Elementary school 1	20	68	44	43			175
Elementary school 2	90	85	45	74			294
Middle school					156	205	361
Total per grade	110	153	89	117	156	205	830

that subjects select, from among four response words, the word that best completed a sentence which had a deletion. Target words were identical within each test level across all three formats. Figure 1 shows sample items from Level 4 of the vocabulary test for each of the three test formats. (Appendix A contains the stimuli for all Level 4 tests.)

All levels of the reading subtest of the Metropolitan Achievement Tests were Form F, and were administered in accordance with directions specified in the Teacher's Directions manuals.

Procedure

Two tests were administered to all subjects participating in the study: one of the three formats of the vocabulary test, and the appropriate level of the reading subtest of the Metropolitan Achievement Tests (see Table 2). Second through fifth grade subjects received the reading subtest of the Metropolitan Achievement Tests first.

All testing took place in the schools and was conducted by members of the Project staff. Classroom teachers did not assist with the administration of tests. Second through fifth grade subjects were tested in their classrooms. Sixth grade subjects were tested in the middle school's reading laboratory, a large room which was partitioned into three smaller rooms. Seventh grade subjects were tested in three different areas within the middle school's Instructional Media Center (IMC).

Each class or group of subjects was assigned to one of the three vocabulary test formats. Each subject was tested on three consecutive levels of the assigned format: a test one grade below the subject's

Synonym Test (Level 4)

2.		gain		
	need	have	put on	lose
3.		nervous		
	hungry	uneasy	young	tired

Synonym in Context Test (Level 4)

2.	Ann will gain weight if she eats that for lunch every day.			
	need	have	put on	lose
3.	Patty bit her nails because she was nervous.			
	hungry	uneasy	young	tired

Cloze Test (Level 4)

2.	Ann will _____ weight from eating too much.			
	need	have	gain	lose
3.	Patty bit her nails becuase she was _____.			
	smart	nervous	young	asleep

Figure 1. Sample items from level 4 of the vocabulary test.

Table 2

Subjects by Grade Level, Format and Level of Vocabulary Test, and Level of the Metropolitan Achievement Tests

Grade level	Number of subjects	Format and levels of vocabulary tests	Level of Metropolitan Achievement Tests (Form F)
2	21	Synonym (levels 1, 2, 3)	Primary II
2	43	Synonym in Context (levels 1, 2, 3)	Primary II
2	41	Cloze (levels 1, 2, 3)	Primary II
3	60	Synonym (levels 2, 3, 4)	Elementary
3	47	Synonym in Context (levels 2, 3, 4)	Elementary
3	44	Cloze (levels 2, 3, 4)	Elementary
4	22	Synonym (levels 3, 4, 5)	Elementary
4	22	Synonym in Context (levels 3, 4, 5)	Elementary
4	52	Cloze (levels 3, 4, 5)	Elementary
5	47	Synonym (levels 4, 5, 6)	Intermediate
5	45	Synonym in Context (levels 4, 5, 6)	Intermediate
5	18	Cloze (levels 4, 5, 6)	Intermediate
6	27	Synonym (levels 5, 6, 7)	Intermediate
6	20	Synonym in Context (levels 5, 6, 7)	Intermediate
6	56	Cloze (levels 5, 6, 7)	Intermediate

Table 2 (continued)

Grade level	Number of subjects	Format and levels of vocabulary tests	Level of Metropolitan Achievement Tests (Form F)
7	44	Synonym (levels 6, 7, 8)	Advanced
7	5	Synonym in Context (levels 6, 7, 8)	Advanced
7	156	Cloze (levels 6, 7, 8)	Advanced

grade level. Overall, subjects were tested on 30 items over 3 grade levels of the assigned vocabulary test format.

Prior to beginning the vocabulary test, pupils were instructed to look at an item on the chalkboard. The chalkboard item was an example of the type of items that would appear on the test within the assigned vocabulary test format. The stimulus word or stimulus sentence and the four response words, were read aloud by the examiner and pupils together. Pupils were asked to select the correct response word from the four choices.

Following the chalkboard item, pupils read the directions printed in their test booklets and did the two practice items beneath the directions. After the Project staff member discussed the two practice items, pupils were given 5 minutes to work independently on the first level of the vocabulary test. The examiner then rephrased the test directions, and pupils were told to proceed with the next level of the vocabulary test. As with the earlier level, pupils were allowed 5 minutes for test completion. The third level of the vocabulary test was conducted in the same way as the two preceeding levels. When this last level of the test was completed, the vocabulary test booklets were collected.

Total testing time for all subjects in grades two through seven, on both the vocabulary test and the reading subtest of the Metropolitan Achievement Tests, ranged from 60 to 75 minutes, with the longer times being taken by younger students.

Results

Table 3 presents summary data for each of the three vocabulary assessment formats tested. It should be noted that the mean scores shown for each format were calculated on all eight levels of the tests across the six grades tested in the study. Each subject received three levels of a particular format of the vocabulary test; one just below, one at, and one just above his or her grade level.

Table 4 shows the correlations between the three vocabulary formats and the reading subtest of the Metropolitan Achievement Tests using Pearson Product moment correlations.

Examination of Table 4 indicates no markedly high relationship for any particular format, either across or within grades. Correlations of each format with the comprehension measure, however, were significant at the .001 level in each case, with the exception of the synonym in context format for seventh grade subjects. Here the small cell size (n = 5, p = .118) precluded meaningful correlation figures. Due to an error in collation of the test booklets, most of the data for this cell could not be used.

Table 5 shows Pearson correlations formats collapsed across grades and test levels. The synonym in context format achieved the highest overall correlation with comprehension although all three formats correlated significantly (at or above .001) with the criterion measure.

Table 6 presents the results of t-tests which were performed on scores under the three vocabulary formats. The synonym/synonym in

Table 3

Raw Score Summary Statistics

on Vocabulary Tests

Formats	N	$\bar{X}$	SD	SE	Range of 30 possible items
Synonym	221	19.98	5.31	.35	2-30
Synonym in Context	182	20.95	5.70	.42	2-30
Cloze	367	20.88	5.65	.29	3-30
Total Group	770	20.63	5.57	.20	2-30

Table 4

Pearson Correlations of Vocabulary

Formats x Metropolitan Subtest

Listed by Grade

Grade	Format		
	Synonym	Synonym in Context	Cloze
2	.7341	.6982	.8088
3	.7746	.7711	.7011
4	.7377	.6360	.6523
5	.6925	.7287	.8584
6	.6455	.7446	.7914
7	.7396	.6495[a]	.6646

[a]Cell size = 5, p = .118, all others significant at .001 level.

Table 5

Pearson Correlations of Vocabulary

Formats x Metropolitan Subtest

Across Grades

	Synonym	Synonym in Context	Cloze
r	.6730	.6884	.6410

Table 6

t-test Comparisons of Vocabulary

Formats (Unstratified on Metropolitan Subtest)

Comparisons	$\bar{X}$ Difference	*t*-value	Significance probability
Synonym/Synonym in Context	1.38	4.093	.00001
Synonym/Cloze	.73	2.534	.0115
Cloze/Synonym in Context	-.65	-2.139	.0327

context comparison yielded a t-value of 4.093, significant at the .00001 level. The synonym/cloze comparison also achieved significance with a t-value of 2.534 at the .0115 level. The cloze/synonym in context comparison fell short of significance at the .01 level with an absolute value for t of 2.139 at the .0327 level.

Table 7 presents the results of a one-way analysis of variance which was performed to examine the main effect of format. This is indicated on the table under the heading "between groups." The F-ratio for the format effect was 2.190 with a probability of .110. Because this F-value may have been lowered somewhat by the wide variation among subjects' reading comprehension abilities within format groups, it was decided that subjects' scores should be stratified into three comprehension level groups based on their performance on the criterion measure (the reading subtest of the Metropolitan Achievement Tests). Analysis on the stratified groups was then performed to determine whether the three comprehension groups performed differentially on the three vocabulary formats.

On the basis of a cumulative frequency table of subjects' raw scores on the reading subtest of the Metropolitan Achievement Tests, cut-off points were established which divided subjects into low, medium, or high comprehension levels. The comprehension ability groupings which resulted from this stratification procedure had the characteristics shown in Table 8.

Table 9 presents summary statistics on the nine comprehension levels x format subgroupings.

Table 7

One Way Anova of

Main Effect of Format

Source	DF	SS	MS	F-Ratio	F-Probability
Between groups	2	135.6758	67.8379	2.190	.110
Within groups	772	23917.9883	30.9818		
Total	774	24053.6641			

Table 8

Summary Statistics of

Comprehension Ability Groups

Comprehension level	N	Percent of total N	Range on Metropolitan Subtest (of 44 possible)
High	238	28.44	34-44
Medium	313	40.65	23-33
Low	219	28.44	3-22

Table 9

Summary Statistics of Format Groups

Stratified into Comprehension Ability Groups

Format Groups		N	Range on Metropolitan Subtest (of 44 possible)	$\bar{X}$	Variance	SD	SE of $\bar{X}$
	Low	55	7-22	17.53	17.40	4.17	.562
Synonym	Medium	89	23-33	27.70	10.12	3.18	.337
	High	77	34-42	37.55	5.49	2.34	.267
Synonym	Low	58	3-22	16.48	17.34	4.16	.547
in	Medium	72	23-33	28.06	8.45	2.91	.343
Context	High	52	34-44	38.04	9.68	3.11	.432
	Low	106	6-22	16.80	20.50	4.53	.440
Cloze	Medium	152	23-33	28.53	9.48	3.08	.250
	High	109	34-43	37.72	6.39	2.53	.242
Total		770	3-44	27.93	75.72	8.70	.314

All-ways comparisons were next made via t-tests. Results of these analyses are summarized in Table 10. As was expected, all comparisons both within and between formats between high and low comprehension levels were significant at or above the .00001 level.

Table 11 presents the within-format comparisons of low versus medium comprehenders and of medium versus high comprehenders. Again, all t-values were uniformly high and significant at or above the .00001 level.

The comparisons of real interest, however, were those of same-comprehension level (SCL) subjects whose vocabulary knowledge was assessed by different formats. Results of these t-test comparisons are presented in Table 12. Two comparisons did show interesting differences, although no t-values were significant at the .05 level. Low comprehenders who took the synonym format performed somewhat better (mean difference = 1.04) than low comprehenders who took the synonym in context format. The t-value for this comparison was 1.65 at a probability level of .0993. The other comparison of note was that between medium comprehenders taking the synonym format with medium comprehenders taking the cloze format of the vocabulary test. The mean difference between these two SCL groups was 0.83 with a t-value of 1.85 with a probability of .0649. All other SCL comparisons had t-values of 1.3 or less and significance levels of 0.195 or more.

The final analysis of the data consisted of sequential tests via two-way analysis of variance under first Fixed and then Random Effects models. Table 13 presents the results of these tests.

Table 10

t-tests of Unweighted Means of Factors of the Metropolitan Subtest and Vocabulary on Stratified Groups

Stratified groups	Synonym/ Low	Synonym/ Medium	Synonym/ High	Synonym in Context/Low	Synonym in Context/Medium	Synonym in Context/High	Cloze/Low	Cloze/ Medium	Cloze High
Synonym/Low	-----								
Synonym/ Medium	t=17.63 p>.0001	-----							
Synonym/ High	t=36.03 p>.0001	t=18.82 p>.0001	-----						
Synonym in Context/Low	t=1.65 p=.0093	t=19.76 p>.0001	t=36.03 p>.0001	-----					
Synonym in Context/Medium	t=17.48 p>.0001	t= .673 p=.5009	t=17.22 p>.0001	t=19.51 p>.0001	-----				
Synonym in Context/High	t=31.54 p>.0001	t=17.62 p>.0001	t= .817 p=.4143	t=33.58 p>.0001	t=16.31 p>.0001	-----			
Cloze/Low	t=1.30 p>.1946	t=22.54 p>.0001	t=41.20 p>.0001	t= .581 p=.5614	t=21.92 p>.0001	t=37.30 p>.0001	-----		
Cloze/ Medium	t=20.79 p>.0001	t=1.85 p=.0649	t=19.18 p>.0001	t=23.21 p>.0001	t= .979 p=.3281	t=17.61 p>.0001	t=27.55 p>.0001	-----	
Cloze/High	t=36.32 p>.0001	t=20.88 p>.0001	t= .358 p=.7203	t=38.87 p>.0001	t=18.94 p>.0001	t= .554 p=.5801	t=45.62 p>.0001	t=21.80 p>.0001	-----

Table 11

Comparison of Comprehension Groups

Within Formats of Vocabulary Measure

Formats	Low :: Medium	Medium :: High
Synonym	t=17.63	t=18.82
Synonym in Context	t=19.51	t=16.31
Cloze	t=27.55	t=21.80

Note. All t-values significant at or above .00001 level.

Table 12

Comparison of Comprehension Ability Groupings Across Vocabulary Formats

Formats	Low	Medium	High
Synonym/Synonym in Context	t=1.65 (.0993)	t=.673 (.5009)	t=.817 (.4143)
Synonym/Cloze	t=1.30 (.1946)	t=1.85 (.0649)	t=.358 (.7203)
Synonym in Context/ Cloze	t=.581 (.5614)	t=.979 (.3281)	t=.554 (.5801)

Note. Significance levels shown in parentheses.

Table 13

Sequential Hypotheses Test Via Two Way Anova

Under Fixed and Random Effects Models

Source	df	Dependent variable	F-ratio	Probability level
		Fixed Effects Model		
Format	$\frac{2}{761}$	Metropolitan	.3842	.6811
		Vocabulary	6.3831	.0018
Metropolitan Subtest	$\frac{2}{761}$	Metropolitan	2191.13	.0000
		Vocabulary	222.96	.0000
Interaction	$\frac{4}{761}$	Metropolitan	1.5869	.1758
		Vocabulary	.3567	.8394
		Random Effects Model		
Format	$\frac{2}{4}$	Metropolitan	.2421	.7957
		Vocabulary	17.8933	.0101
Metropolitan Subtest	$\frac{2}{761}$	Metropolitan	2191.13	.0000
		Vocabulary	222.960	.0000
Interaction	$\frac{4}{761}$	Metropolitan	1.5869	.1758
		Vocabulary	.3567	.8394

The format effect on the vocabulary measures was significant at the .01 level with F-ratios of 6.38 and 17.89 under the Fixed and Random Effects models, respectively. No interaction effects were present in either case.

Summary

Subjects performed similarly on the three assessment formats across the six grades tested. No one test format performed remarkably better than either of the others within or across grades. But, there were significant mean differences in scores obtained under the three formats overall. These differences indicated synonym, synonym in context, and cloze in rank order of difficulty, from easiest to most difficult.

When subjects were stratified by comprehension ability, once again scores on any one format were not significantly higher overall, than scores on either of the other two formats. Yet, some differences did appear. Low comprehenders scored slightly higher on synonym than on synonym in context, and medium comprehenders scored somewhat higher on cloze than on synonym in context.

These findings are in contrast to the results presented in the PLATO IV report (Johnson et al., 1978). In the studies described in that report, on-line testing of the same three vocabulary formats was conducted, and the reading subtest of the Metropolitan Achievement Tests was also administered. In these studies, however, fourth and fifth grade scores showed far higher correlations between vocabulary and comprehension scores than did second and third grade scores. The overall results were similar in one important respect: no one format of

vocabulary assessment showed consistently higher correlations across the grade levels tested.

It is possible that further differences in vocabulary performance might appear if subjects were grouped into finer categories of comprehension level. Such elaborate stratifications were not believed to be justified at this point, because investigators questioned the basis for making fine discriminations among performances based on a single standardized measure of comprehension. But, future studies of format effects in vocabulary assessment might involve other, concurrent measures of comprehension ability. Ultimately, the validity and nature of the format effects discovered in the study described above rest on the validity and reliability of the criterion measure of comprehension used. As more effective comprehension assessment instruments become available, more confidence can be placed in the results of studies on format effects which utilize those comprehension assessment instruments as criterion measures.

Although the findings of this study did not demonstrate the superiority of one vocabulary paradigm over another, some interesting format effects did appear. Further examination of the question of format effects on the assessment of vocabulary knowledge is certainly warranted. If a particular format used to assess vocabulary knowledge yields lower scores than another format, this information should be made available to educators utilizing those scores. It may well be that no one paradigm for assessing vocabulary is most effective across all grades and ability groups of students. Nevertheless, information

is needed about specific assessment modes, so that informed choices can be made about which of these modes are appropriate for particular testing purposes.

References

Block, K. K. Vocabulary development: A problem in learning and assessment. Unpublished manuscript, Pittsburg: Learning Research and Development Center, 1976.

Carroll, J. B. Words, meanings and concepts. Harvard Educational Review, 1964, 34, 178-202.

Dale, E. Bibliography of vocabulary studies (5th ed.). Columbus, Ohio: Ohio State University, 1965.

Dale, E., & O'Rourke, J. The living word vocabulary. Elgin, Ill.: Dome, Inc., 1965.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Johnson, D. D., Pittelman, S., Schwenker, J., & Perry, J. On line diagnosis of reading difficulties - IV (Technical Report No. 464). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1978.

Otto, W., & Chester, R. Sight words for beginning readers. The Journal for Educational Research, 1972, 65 (10), 435-443.

Spearritt, D. Identification of subskills of reading comprehension by maximum likelihood factor analysis, Reading Research Quarterly, 1972, 8, 92-111.

Venezky, R., Perry, J., Chicone, S. & Pittelman, S. Summary of studies for an on-line reading diagnosis system - III (Technical Report No. 409). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1977.

Appendix A

Stimuli from Level 4 of the Vocabulary Tests

Synonym Test, Level 4

Synonym in Context Test, Level 4

Cloze Test, Level 4

Synonym Test

Level 4

Stimulus	Responses			
gigantic	huge	baby	friendly	lonely
gain	need	have	put on	lose
nervous	hungry	uneasy	young	tired
broken	destroyed	shaken	dirtied	moved
concealed	wanted	discovered	found	hid
feeble	funny	unhappy	weak	strong
briskly	slowly	quickly	sadly	awkwardly
exhibits	acts	people	rides	displays
surprised	sick	gone	bored	amazed
twisted	held	wound	found	placed

Synonym in Context Test

Level 4

Stimulus (The target word is underlined.)	Responses	
The gigantic giraffe licked the child's hand.	huge	baby
	friendly	lonely
Ann will gain weight if she eats that for lunch everyday.	need	have
	put on	lose
Patty bit her nails because she was nervous.	hungry	uneasy
	young	tired
Many of our dishes were broken by the earthquake.	destroyed	shaken
	dirtied	moved
He concealed the gun in his coat pocket.	wanted	discovered
	found	hid
The feeble old dog walked beside the woman.	funny	unhappy
	weak	strong
The horse trotted briskly along the path.	slowly	quickly
	sadly	awkwardly
There were many interesting exhibits at the fair.	acts	people
	rides	displays

Stimulus (The target word is underlined.)	Responses	
Mark was <u>surprised</u> when his friend from Texas arrived.	sick	gone
	bored	amazed
Cathy <u>twisted</u> a rubber band around her finger.	held	wound
	found	placed

Cloze Test

Level 4

Stimulus	Responses	
The giraffe was so _____ that he made even the elephants look small.	gigantic	young
	friendly	lonely
Ann will _____ weight from eating too much.	need	have
	gain	lose
Patty bit her nails because she was _____.	smart	nervous
	young	asleep
Many of our dishes were _____ by the earthquake and we had to sweep up all the pieces.	broken	shaken
	lost	moved
He _____ the gun in his coat pocket so no one would know he had it.	made	discovered
	found	concealed
The _____ old dog could barely keep up with the woman, but wagged his tail happily anyway.	quick	unhappy
	feeble	strong
The horse trotted _____ along the path and arrived home sooner than usual.	slowly	briskly
	sadly	awkwardly
There were many interesting _____ at the science fair to show how different foods are grown.	acts	people
	rides	exhibits

Stimulus	Responses	
Mark was _____ and happy when his friend arrived unexpectedly from Texas.	sick	gone
	bored	surprised
.thy _____ around so she could reach he book on the shelf behind her.	held	twisted
	found	placed

Center Planning and Policy Committee

Richard A. Rossmiller
Wayne Otto
Center Co-Directors

Dale D. Johnson
Area Chairperson
Studies in Language:
Reading and Communication

Marvin J. Fruth
Area Chairperson
Studies of Implementation
of Individualized Schooling

Penelope L. Peterson
Area Chairperson
Studies of Instructional Programming
for the Individual Student

James M. Lipham
Area Chairperson
Studies of Administration and
Organization for Instruction

Thomas A. Romberg
Area Chairperson
Studies in Mathematics and Evaluation
of Practices in Individualized Schooling

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Associate Professor
Curriculum and Instruction

Louise J. Cherry
Assistant Professor
Educational Psychology

William J. Davis
Assistant Professor
Educational Administration

W. Patrick Dickson
Assistant Professor
Child and Family Studies

Lloyd E. Frohreich
Associate Professor
Educational Administration

Marvin J. Fruth
Professor
Educational Administration

Dale D. Johnson
Professor
Curriculum and Instruction

Herbert J. Klausmeier
V.A.C. Henmon Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

James M. Lipham
Professor
Educational Administration

Dominic W. Massaro
Professor
Psychology

Donald M. McIsaac
Professor
Educational Administration

Wayne Otto
Professor
Curriculum and Instruction

Penelope L. Peterson
Assistant Professor
Educational Psychology

Thomas S. Popkewitz
Associate Professor
Curriculum and Instruction

Gary G. Price
Assistant Professor
Curriculum and Instruction

W. Charles Read
Associate Professor
English and Linguistics

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Peter A. Schreiber
Associate Professor
English and Linguistics

B. Robert Tabachnick
Professor
Curriculum and Instruction

Gary G. Wehlage
Associate Professor
Curriculum and Instruction

Steven R. Yussen
Associate Professor
Educational Psychology

9/78